BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e37525
doi: 10.3897/biss.3.37525

OPEN ACCESS

Conference Abstract

Discovering Patterns of Biodiversity in Insects Using Deep Machine Learning

Chandra Earl[‡], Alexander E. White[§], Michael G. Trizna[§], Paul B. Frandsen[|], Akito Y. Kawahara[¶], Sean G. Brady[§], Rebecca B. Dikow[§]

‡ University of Florida, Gainesville, Fl, United States of America
§ Smithsonian Institution, Washington DC, United States of America
| Brigham Young University, Provo, Utah, United States of America
¶ Florida Museum of Natural History, Gainesville, United States of America

Corresponding author: Rebecca B. Dikow (dikowr@si.edu)

Received: 19 Jun 2019 | Published: 02 Jul 2019

Citation: Earl C, White A, Trizna M, Frandsen P, Kawahara A, Brady S, Dikow R (2019) Discovering Patterns of Biodiversity in Insects Using Deep Machine Learning. Biodiversity Information Science and Standards 3: e37525. https://doi.org/10.3897/biss.3.37525

Abstract

Museum specimens have enormous potential for use in a broad range of biodiversity and evolutionary questions, but their data are typically accessible only to researchers who can physically visit collections facilities. Recent digitization efforts of collections provide new modes of access and collaboration to enrich biodiversity knowledge, and remarkable progress is now being made in assembling a corpus of imaged specimens and their associated labels. The Smithsonian Digitization Program Office recently partnered with the National Museum of Natural History (NMNH), Department of Entomology to mass-digitize their bumblebee (genus *Bombus*) collection. Digital images were captured from more than 45,000 specimens and labels were transcribed by volunteers through the Smithsonian Transcription Center. More than 10,000 of these specimens are not yet identified to subgenus or species. We present deep learning models (specifically, convolutional neural networks) that can classify specimens to subgenus (NMMH has 15 subgenera) and species (NMNH has 178 species). Both models average greater than 90% accuracy even when trained on a small number of input images (tens of images per class). Beyond taxonomic classification, we explore how we can link our models to traditional morphological characters, biogeographical data, digitized scientific literature, and external image datasets to further our understanding of biodiversity.

This is an open access article distributed under the terms of the CC0 Public Domain Dedication.

Keywords

machine learning, biodiversity, insects

Presenting author

Rebecca B. Dikow

Presented at

Biodiversity_Next 2019